

ABSTRACT OF THE DISCLOSURE

A technique for increased efficiency of content delivery over a network is described. Instances of web pages are divided into (1) templates including those elements of a web page that are relatively unchanging and (2) delta information including those elements that are ephemeral or customized. Each template is compressed and cached at an originating server. Transmission of the delta information is decoupled from transmission of the template. When a user requests a page, the compressed template is sent (either from an originating server or a mirror thereof). The delta information is compressed and sent separately. Since the template is only compressed once and is cached locally, it requires less bandwidth and allocation of other computing resources to transmit than transmission of a compressed web page.